

気象情報と Tweet データの統合的分析による体感気温の定量化とその需要予測への利用に関する研究

情報数理応用研究

5215C017-1 馬賀嵩士
指導教員 後藤正幸

Quantification of Sensible Temperature by Analyzing Meteorological Information and Tweet Data and Its Application to Demand Forecasting

MAGA Takashi

1 研究背景・目的

多様な商品を所在地域の異なる多店舗で販売する小売チェーンにおいて、気象条件の変化に伴う需要変動に起因する在庫過多や廃棄処分が課題となっている。これに対し、蓄積が可能となった多様な大規模データを分析することで、ミクロな観点からの需要予測の精度向上とより細かな在庫管理が期待されている。本研究では特に、あるカテゴリの食料品の需要は、客観的な気温よりも消費者が感じる主観的な体感気温に直接的に大きな影響を受けるのではないかとという仮説に着目する。

体感気温の定量化に関する試みは数多くなされてきており、湿度や風速といった気象条件に影響を受けやすいといったことがわかっている。この点に着目した代表的な算出式にミスナールの式、不快指数などがある [1],[2]。加えて、体感気温は湿度や風速、日射量といった気象条件以外にも、着衣量、代謝量のような人体条件の影響も受けるため、その感覚の定量化の方法はさまざまである。

これに対し本研究では、Twitter に投稿される Tweet データのうち、温度感覚にかかわるつぶやきを利用することで、人々がその時々感じている体感気温を定量化することを考える。Twitter は、日本国内のユーザー数が約 3000 万人、1 日の Tweet 数は数億件にも及ぶ大規模 SNS である。数千万人の人々の感覚がテキストの形でリアルタイムに蓄積されるこの Tweet データを利用することで、人々が感じている体感気温をより忠実に定量化できるのではないかと考えられる。

そこで本研究では、人間が感じる体感気温は、Twitter における「暑い」や「寒い」といったつぶやきの形で現れるのではないかとという仮定のもと、大規模テキストデータである Twitter のデータから、体感気温の定量化を行うモデルを提案する。さらに、大手小売チェーンの売上データを用いて、提案モデルにより得られる体感気温と商品需要との関係を分析し、需要予測への適用可能性について検討を行うとともに、体感気温で定義される「暑い日」「寒い日」に売れ行きが伸びる商品カテゴリの抽出を試みる。

2 データ概要・事前分析

本研究において用いるデータは、位置情報付き日本語 Tweet データ (1/10 サンプル)、気象データ (アメダス東京地点)、某小売チェーンのある店舗における商品カテゴリ毎の日別売上数量データ、および気象データ (店舗所在地域) の 4 点である。Tweet データは、2012 年 9 月 1 日から 2015 年 9 月 30 日のもので、各 Tweet には、つぶやかれた日時、位置 (緯度と経度) が付与されており、Tweet 総数は 15,495,108 件である。気象データの期間は、Tweet データと同様であり、第 t 日 (日付) の平均気温 T_t [°C]、最低気温 T_t^{\min} [°C]、最高気温 T_t^{\max} [°C]、相対湿度 H_t [%]、平均風速 W_t [m/s]、日合計降水量 RF_t [mm]、日最深積雪 SF_t [cm]、日照時間 ST_t [時間]、日合計全天日射量 SR_t [MJ/m²] を利用する。売上数量データは、当該店舗で販売されている全 153 カテゴリ (中分類) の 2013

年 1 月 1 日から 2014 年 12 月 31 日における日別売上数量が蓄積されたものである。また、このデータには同期間における店舗所在地域の気象データが付随しており、平均気温、相対湿度、平均風速、日合計降水量、日合計全天日射量の 5 つの観測値が含まれている。本研究では、4 節の体感気温の算出においては Tweet データおよび気象データ (アメダス東京地点) を、5 節の体感気温と需要予測の関係の分析においては、日別売上数量データおよび気象データ (店舗所在地域) を用いるものとする。

ここで「暑い」、「寒い」という単語を含む Tweet を温度感覚の定量化に利用することの妥当性を検証するために、全 Tweet データに対して形態素解析を行い、基本形が「暑い」、「寒い」である単語を含む Tweet の総 Tweet 数に占める割合を日ごとに集計した。これらの Tweet の割合と平均気温の推移を図 1 に示す。

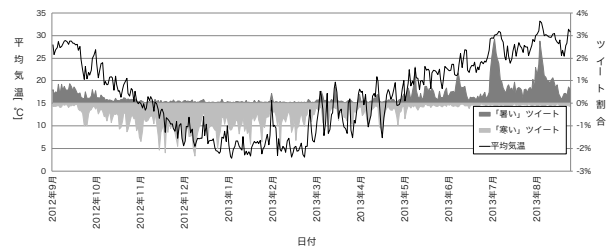


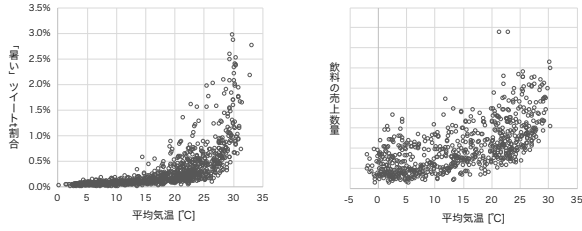
図 1. 「暑い/寒い」 Tweet の割合と平均気温の推移

図 1 より、「暑い/寒い」を含む Tweet の割合と平均気温の間には一定の相関があり、これらの値を温度感覚の定量化に用いることの妥当性が明らかとなった。一方で、平均気温と「暑い/寒い」を含む Tweet の割合の関係に大きくずれが生じている時期も存在しており、これらの時期においては体感気温と平均気温の間に乖離が生じている可能性があることがわかる。

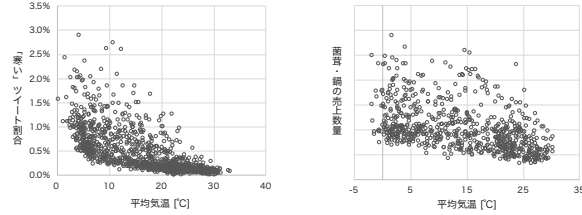
次に、「暑い/寒い」を含む Tweet 割合と商品需要との関係を確認するため、平均気温と「暑い/寒い」を含む Tweet の割合、および平均気温と商品需要の関係を図 2 に示す。なお、本研究では気象感応度が特に高い商品カテゴリを分析対象とするものとし、暑さに影響を受けやすい商品カテゴリとして飲料を、寒さに影響を受けやすい商品カテゴリとして菌茸・鍋を取り上げた。

図 2-(a) より、「暑い」を含む Tweet の割合は、実平均気温に対して曲線的に増加しており、飲料の売上数量も似た傾向を示していることがわかる。また、飲料の売上数量では、平均気温が 15°C 付近において小さな山ができており、この付近の体感の違いが飲料の需要の変動に影響を及ぼしていると考えられる。次に、図 2-(b) を見ると、「寒い」を含む Tweet の割合が平均気温に対して曲線的に減少しているが、「暑い」 Tweet の場合よりも勾配は緩やかであり、菌茸・鍋の売上数量もそれと似た傾向となっていることがわかる。このように、気温に応じた暑さと寒さの感じ方の傾向には違いが見られることから、本

研究においては両者を独立に扱い、暑さ、寒さそれぞれに基づく体感気温を定量化するものとする。



(a) 「暑い」 Tweet 割合と飲料の売上数量



(b) 「寒い」 Tweet 割合と菌茸・鍋の売上数量

図 2. 平均気温と Tweet 割合, 売上数量の関係の比較

3 Tweet データを用いた体感気温の定義

3.1 着想

Twitter を利用しているユーザは、その日に感じた暑さ、寒さに基づいて、“今日はかなり暑い”、“寒すぎて布団から出られない”、のように、「暑い」や「寒い」といった単語を含む形で自身の感じている温度感覚をつぶやいていると考えられる。また、暑さ、寒さの程度に比例して、暑い、寒いと感じるユーザの数は増加し、このような Tweet の数も増加すると考えられるため、これらの Tweet の数が暑さ、寒さの程度を示していると考えことは妥当であるといえる。よって、これらの値と実平均気温との関係をモデル化することで、体感気温の定量化が可能となるのではないかと考えられる。ただし、ここでは「暑い/寒い」を含む Tweet は絶対数ではなく、その日の総 Tweet 数に占めるそれぞれの Tweet の割合を用いるものとする。いま、第 t 日の総 Tweet 数に占める「暑い」を含む Tweet の割合を r_t^{hot} 、「寒い」を含む Tweet の割合を r_t^{cold} 、実平均気温を T_t [°C] とする。本研究では、「暑い/寒い」の Tweet 割合と実平均気温との関係の定量化に回帰モデル [3] を用いるものとする。具体的には、 T_t を説明変数、 r_t^{hot} を目的変数とした回帰分析を行うことで、偏回帰係数 $\hat{\alpha}_0, \hat{\alpha}_1$ で定められ、その日の気温から「暑い」の Tweet 割合を予測する $\hat{r}_t^{hot} = f(T_t | \hat{\alpha}_0, \hat{\alpha}_1)$ なる回帰式が得られる。一方で、この逆関数 f^{-1} は、「暑い」 Tweet 割合からその値に対応する気温を算出する関数 $g(r_t^{hot} | \hat{\alpha}_0, \hat{\alpha}_1) = f^{-1}(r_t^{hot} | \hat{\alpha}_0, \hat{\alpha}_1)$ であり、 r_t^{hot} を含むものもできる。すなわち、この関数 g により、「暑い」 Tweet 割合の値を平均的に生みだし得る気温 S_t^{hot} [°C] を推定することができる。本研究では、この S_t^{hot} を暑さに基づく第 t 日の体感気温と定義する。

ここで、上記の回帰における残差 $\delta_t = r_t^{hot} - \hat{r}_t^{hot}$ は、一定の誤差を含みながら、湿度や風速といった気温以外の気象要素の影響により生じていると仮定し、この δ_t をそれらの気象要素から何らかのモデルで回帰することを考える。回帰による推定値を $\hat{\delta}_t = \delta_t - \varepsilon$ とすると (ε は残差)、 $\hat{r}_t^{hot} = \hat{r}_t^{hot} + \hat{\delta}_t$ は気象要素の影響のみを考慮した場合に想定される第 t 日の「暑い」 Tweet 割合とみなすことができる。この \hat{r}_t^{hot} を先に述べた関数 g に入力することで、気象要素の影響のみを考慮した場合に想定される S_t^{hot} の値を算出することができる。 δ_t から S_t^{hot} を算出することも可能だが、その場合、 S_t^{hot} の算出に r_t^{hot} の値が必要となるため、Tweet データを蓄積する手間を考

えると実用上好ましくない。加えて、Tweet データの性質上、 r_t^{hot} の値には一定のノイズが含まれると考えられるため、 δ_t の値を直接 S_t^{hot} の算出に用いるのは適切ではないといえる。そこで、気象要素から δ_t を推定し、そのときの推定誤差は気象要素以外の要因に起因するノイズであると考え、この問題を解消する狙いもある。 r_t^{cold} についても同様の流れを踏むことで、寒さに基づく第 t 日の体感気温 S_t^{cold} を算出することができる。また、これ以降、 S_t^{hot} と S_t^{cold} を合わせて体感気温と表現することとする。 r_t^{hot} による S_t^{hot} の算出のイメージを図 3 に示す。

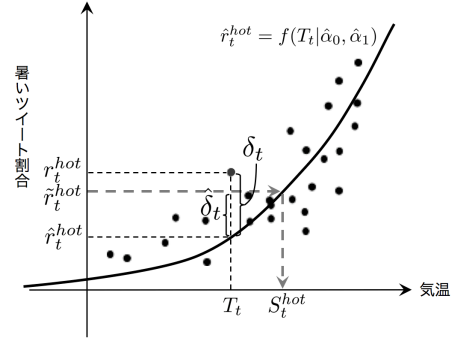


図 3. 体感気温算出のイメージ

3.2 体感気温の定量化

以下では、具体的な S_t^{hot}, S_t^{cold} の算出方法について述べる。まず、平均気温と Tweet 割合との関係を定式化するため、式 (1) の通り回帰式を定義する。

$$R_t^{hot} = \alpha_0^h + \alpha_1^h T_t + \varepsilon_t \quad (1)$$

$$\varepsilon_t \sim \mathcal{N}(0, \sigma_h^2) \quad (2)$$

ただし、説明変数を T_t 、目的変数を $R_t^{hot} = \log(r_t^{hot}/(1 - r_t^{hot}))$ とする。また、誤差 ε_t は独立に平均 0、分散 σ_h^2 の正規分布に従うものとする。次に、この回帰における残差 $\Delta_t^{hot} = R_t^{hot} - (\hat{\alpha}_0^h + \hat{\alpha}_1^h T_t)$ を目的変数とし、気象要素ベクトル \mathbf{w}_t を説明変数とした回帰モデルを構築する。ただし、各気象要素の体感気温への影響の度合いは季節ごとに異なるという仮定から、回帰モデルは季節ごとに構築するものとする。 s_t を、第 t 日が春 (3-5 月) であるとき 1、夏 (6-8 月) であるとき 2、秋 (9-11 月) であるとき 3、冬 (12-2 月) であるとき 4 をとる季節変数とすると、季節 k における Δ_t^{hot} の回帰式は以下のように表せる。

$$\hat{\Delta}_t^{hot} = F_k^{hot}(\mathbf{w}_t | \boldsymbol{\theta}_k^{hot}) \quad (3)$$

$$\forall t \in S_k := \{t | s_t = k\} \quad (4)$$

ここで、 $F(\cdot | \boldsymbol{\theta})$ は、パラメータを $\boldsymbol{\theta}$ とする任意の回帰モデルであるとする。さらに、ここまでの手順を $R_t^{cold} = \log(r_t^{cold}/(1 - r_t^{cold}))$ についても同様に行い、 $\hat{\alpha}_0^c, \hat{\alpha}_1^c$ を推定、 F_k^{cold} を学習する。そして、得られた $\hat{\alpha}_1^h, \hat{\alpha}_1^c, F_k^{hot}, F_k^{cold}$ を用いて、式 (1) の逆関数に対応する以下の式 (5), (6) により S_t^{hot}, S_t^{cold} を算出する。

$$\begin{aligned} S_t^{hot} &= \frac{1}{\hat{\alpha}_1^h} \left((\hat{\alpha}_0^h + \hat{\alpha}_1^h T_t) + F_{s_t}^{hot}(\mathbf{w}_t | \boldsymbol{\theta}_k^{hot}) - \hat{\alpha}_0^h \right) \\ &= \frac{1}{\hat{\alpha}_1^h} \left(\hat{\alpha}_1^h T_t + F_{s_t}^{hot}(\mathbf{w}_t | \boldsymbol{\theta}_k^{hot}) \right) \\ &= T_t + \frac{1}{\hat{\alpha}_1^h} F_{s_t}^{hot}(\mathbf{w}_t | \boldsymbol{\theta}_k^{hot}) \end{aligned} \quad (5)$$

$$S_t^{cold} = T_t + \frac{1}{\hat{\alpha}_1^c} F_{s_t}^{cold}(\mathbf{w}_t | \boldsymbol{\theta}_k^{cold}) \quad (6)$$

体感気温の算出アルゴリズムを以下に示す。

[体感気温の算出アルゴリズム]

Step1) 式 (1), (2) で定義される回帰分析を行い, 偏回帰係数 α_0^h, α_1^h を推定する。

Step2) 同様に, R_t^{cold} についても α_0^c, α_1^c を推定する。

Step3) 各季節 ($k = 1 \sim 4$) について, F_k^{hot} を学習する。

Step4) 同様に, R_t^{cold} についても F_k^{cold} を学習する。

Step5) 式 (5), (6) により S_t^{hot}, S_t^{cold} を算出する。 □

4 体感気温の算出

3 節で定義した体感気温の算出方法により, 実際の Tweet データと気象データ (アメダス東京地点) を用いて体感気温の算出モデルを構築し, それを用いて体感気温の算出を行う。対象期間は, 2012 年 9 月 1 日から 2015 年 9 月 30 日であり, 「暑い/寒い」の Tweet 割合に対して, 対移動平均比率法 [4] を適用することで曜日効果の除去を行った。また, 気象要素ベクトルは, 気象データ (店舗所在地) に共通する気象要素のみを含むよう $w_t = (T_t - T_{t-1}, H_t, W_t, RF_t, SR_t)^T$ とし, 回帰モデル F には Random Forest [5] を用いるものとする。Random Forest におけるパラメータは Grid Search により決定した。

各月における実平均気温と体感気温の月平均値の推移, および各月の標準偏差を図 4 に示す。ただし, 図中の $\mu_T, \mu_{Sh}, \mu_{Sc}$ はそれぞれ $T_t, S_t^{hot}, S_t^{cold}$ の各月の平均値を, $\sigma_T, \sigma_{Sh}, \sigma_{Sc}$ は同指標の各月の標準偏差を指すものとする。図 4 より, 1 月~7 月は実平均気温よりも体感気温の方が高く, 8 月~12 月では体感気温の方が低くなっていることがわかる。また, S_t^{hot} に比べて S_t^{cold} の方が体感気温との差が大きくなる傾向にあることがわかる。さらに, 短期間で気温変化が多い 5 月, および下降し始める 10 月, 11 月頃に両者の差が大きくなっており, 急激な気温変化の体感気温への影響が示唆される結果となった。この差を季節単位で見ると, 実平均気温が下がり始める秋 (9~11 月) には平均して両者の差が大きく, 低い気温で安定している冬 (12~2 月) には平均して両者の差が小さいことがわかる。このことから, 人間は「暑い → 寒い」の変化に特に敏感であること, 冬の時期は低い気温に慣れ, 寒さの感じ方があまり変動しないことが考えられる。また, 標準偏差 (図中棒グラフ) に注目すると, 平均気温と比較して, 7 月~8 月において S_t^{hot} の, 9 月~2 月において S_t^{cold} のばらつきが大きくなっていることがわかる。これらは, それぞれ「暑い」「寒い」を含む Tweet が多くなりやすい時期であり, これらの変動に大きく影響を受けたためではないかと考えられる。

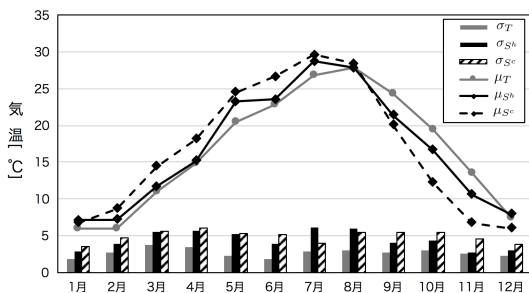


図 4. 月ごとの実平均気温と体感気温の比較

次に, F_k^{hot}, F_k^{cold} の学習の過程で得られた各特微量の重要度を表 1 に示す。なお, 春 ($k = 1$), 夏 ($k = 2$) に関しては F_k^{hot} の値について, 秋 ($k = 3$), 冬 ($k = 4$) に関しては F_k^{cold} の値について考察を行うものとする。春 ($k = 1$) は他の季節と比較して, 日合計全日射量, 平均

風速が体感気温に大きな影響を及ぼしており, 強い風が体感気温に大きな影響を及ぼしていると考えられる。夏 ($k = 2$) は日合計全日射量, 相対湿度が大きな影響を与えており, 日差しが強く蒸し暑い夏の特徴と一致している。秋 ($k = 3$) は相対湿度, 平均風速の影響が大きく, 図 4 から明らかなように, この時期は体感気温が実平均気温に比べて低くなっていることから, 湿度の低下に伴う乾燥と強い風が体感気温の低下を引き起こしていると考えられる。冬 ($k = 4$) は特に日合計全日射量の影響が大きく, 冬の温度感覚は主に日射量に影響を受けていると考えられる。

表 1. 各特微量の重要度

		$k = 1$	$k = 2$	$k = 3$	$k = 4$
F_k^{hot}	気温前日差	0.165	0.176	0.217	0.231
	相対湿度	0.197	0.201	0.202	0.208
	平均風速	0.207	0.183	0.175	0.188
	日合計降水量	0.153	0.116	0.172	0.103
	日合計全日射量	0.278	0.324	0.234	0.271
F_k^{cold}	気温前日差	0.265	0.264	0.194	0.227
	相対湿度	0.216	0.180	0.296	0.204
	平均風速	0.203	0.197	0.210	0.218
	日合計降水量	0.091	0.114	0.102	0.031
	日合計全日射量	0.224	0.245	0.198	0.320

5 体感気温と商品需要の関係に関する分析

本節では, 体感気温と商品需要の関係について分析を行う。5.1 節では, 実平均気温, 体感気温それぞれを用いて需要予測を行った場合の精度を比較し, 体感気温が実平均気温よりも需要変動を説明できているかを検証する。5.2 節では, 体感気温と実平均気温の関係に基づいて「暑い日」「寒い日」を定義し, それぞれにおいて売れ行きが伸びる商品カテゴリの特定を試みる。この結果から, 特定された商品カテゴリに対する細かな在庫管理の一助となる知見が得られることを期待する。

用いるデータは, 2013 年 1 月 1 日から 2014 年 12 月 31 日における, 大手小売チェーンのある店舗の日ごとの各商品カテゴリの売上数量データ, および店舗の所在地の気象データである。同地域の体感気温は, 4 章で得られたパラメータおよび回帰モデルを式 (5), (6) に代入することで得られる体感気温の算出式に, 同地域の気象データを入力することで得られる値を用いるものとする。

5.1 体感気温を用いた需要予測

本節では, 全カテゴリのうち気象感応度が高く, 売上数量が比較的多い菌茸・鍋, こんにゃく, 練り製品, 飲料, 乳飲料・清涼飲料の 5 カテゴリを対象として体感気温を用いた需要予測を行い, 体感気温を用いた場合の結果との比較から体感気温の需要予測への利用可能性を検討する。

5.1.1 実験条件

対象期間のうち, 前半 365 日間の体感気温, および同期間の対象商品カテゴリの売上数量データを学習データ, 後半 365 日間の同データをテストデータとして実験を行った。予測には重回帰モデルを用いることとし, 用いる説明変数は, 平均気温, 体感気温ともに, 気温, 気温の前日差, 前々日差, 曜日変数とする。ここで, 曜日変数とは, 6 つのダミー変数からなるものとし, それぞれ, 土曜日, 日曜日, 月曜日, 火曜日, 水曜日, 木曜日のとき 1, それ以外のとき 0 をとるものとする。精度評価については, 式 (7) で定義される平均絶対誤差率を用いるものとする。なお, N_t は第 t 日の売上数量, \hat{N}_t は N_t の予測値, D を算出対象期間の日数とする。

$$\text{平均絶対誤差率} [\%] = \frac{\sum_{t=1}^D \frac{|\hat{N}_t - N_t|}{N_t}}{D} \times 100 \quad (7)$$

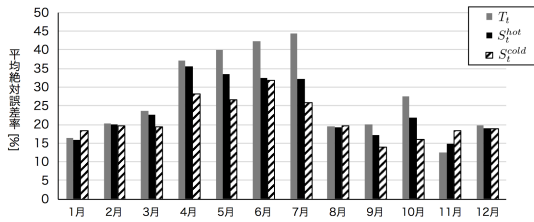
5.1.2 実験結果

実験結果を表2に示す。表2より、菌茸・鍋、こんにゃく、練り製品といった寒さに影響を受けやすいと考えられる商品カテゴリは S_t^{cold} による推定精度が良く、飲料、乳飲料・清涼飲料といった暑さに影響を受けやすいと考えられる商品カテゴリは S_t^{hot} による推定精度が良いことがわかる。

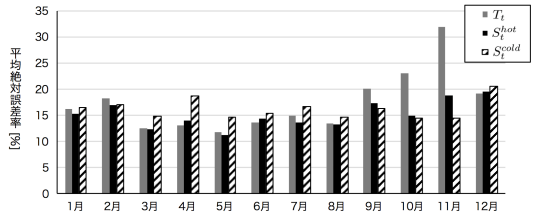
表2. 商品カテゴリごとの平均絶対誤差率 [%]

	平均気温 T_t	体感気温	
		S_t^{hot}	S_t^{cold}
菌茸・鍋	26.9	23.7	21.4
こんにゃく	30.8	28.5	28.3
練り製品	21.4	20.6	20.2
飲料	17.3	15.1	16.2
乳飲料・清涼飲料	12.7	12.4	12.9

また、菌茸・鍋、飲料について、月ごとの平均絶対誤差率を図5に示す。図5-(a)より、菌茸・鍋では特に4月～7月において、平均気温に比べて S_t^{hot} および S_t^{cold} の予測精度が優れており、得られた体感気温が、春から夏にかけての需要変動の説明に効果的であることがわかった。また、図5-(b)より、飲料では特に9月～11月において、 S_t^{hot} および S_t^{cold} の予測精度が優れており、これらが秋の需要変動を説明できていることがわかった。



(a) 菌茸・鍋



(b) 飲料

図5. 月ごとの平均絶対誤差率の比較

5.2 体感気温と商品の売れ行きの関係の分析

5.2.1 実験条件

本節では、体感気温を用いて、「暑い日」または「寒い日」に売れ行きが伸びる商品カテゴリの特定を行う。ここで、商品需要は実平均気温と比較したときの体感気温の高低差に影響を受けやすいのではないかとという仮説に基づき、この高低差と商品の売上数量との関係に注目する。ここでは、体感気温と実平均気温の差 $U_t = S_t^i - T_t$ について、一週間移動平均を取り (これを U_t^m とする)、 $\tilde{U}_t = U_t - U_t^m$ の値を算出し、この値に応じて各日 t に以下によりラベル付けを行う。

$$l_t = \begin{cases} 1 (\text{寒い日}) & \text{if } \tilde{U}_t \leq -\sigma_U \\ 2 (\text{通常の日}) & \text{if } -\sigma_U < \tilde{U}_t \leq \sigma_U \\ 3 (\text{暑い日}) & \text{if } \tilde{U}_t > \sigma_U \end{cases} \quad (8)$$

ただし、 σ_U は \tilde{U}_t の標準偏差とする。ここで、商品カテゴリ i の全期間での平均売上数量を \bar{n}_i 、ラベルが l で

ある日の商品カテゴリ i の平均売上数量を $\bar{n}_{i\ell}$ として、商品カテゴリ i の売上数量の伸び率 G_i を以下のように定義する。

$$G_i = \max_{1 \leq \ell \leq 3} \frac{\bar{n}_{i\ell} - \bar{n}_i}{\bar{n}_i} \quad (9)$$

また、 G_i の値をとるときのラベル ℓ を L_i とする。実験においては、全153カテゴリのうち売上数量上位50件の商品カテゴリのみについて G_i を算出するものとする。

5.2.2 実験結果

売上数量伸び率 G_i が上位10件の商品カテゴリ、およびそのときのラベル L_i を表3に示す。

表3. 売上数量伸び率の上位10件

S_t^{hot}			S_t^{cold}		
商品カテゴリ	G_i	L_i	商品カテゴリ	G_i	L_i
菌茸・鍋	0.201	1	土物	0.267	3
飲料	0.118	3	サラダ	0.172	3
季節・その他	0.118	3	季節・その他	0.168	3
冷凍食品	0.097	1	菌茸・鍋	0.160	1
葉物・妻物	0.094	3	葉物・妻物	0.134	3
鮮魚	0.094	1	鮮魚	0.089	1
サラダ	0.068	3	冷凍食品	0.074	1
土物	0.066	3	水物・加工野菜	0.049	3
練り製品	0.056	1	とり肉	0.039	1
卵	0.055	1	焼き菓子	0.039	3

表3を見ると、 S_t^{hot} と S_t^{cold} のどちらを用いた場合にも、菌茸・鍋、季節・その他、葉物・妻物といった商品カテゴリが共通して特に売れ行きが伸びていることがわかる。また、それぞれの商品カテゴリのラベル L_i を見ると、寒い日 ($L_i = 1$) には菌茸・鍋、冷凍食品、鮮魚の売上数量が、暑い日 ($L_i = 3$) には、土物、葉物・妻物、サラダ、季節・その他の売れ行きが伸びる傾向にあることがわかる。これらの商品カテゴリは本研究の成果として得られた「気象感応度の高い商品カテゴリ」であり、これらの商品カテゴリに対して、各日のラベルをもとに $100G_i\%$ の需要増を見込んだ在庫管理を行うことで、機会損失を減らし、当該商品の売上高を増加させると同時に在庫保管費用の低減が可能となるのではないかと考えられる。

6 まとめと今後の課題

本研究では、Tweet データを用いた体感気温の定量化法を提案した。また、算出した体感気温と商品需要との関係进行分析し、需要予測への利用可能性の検討、および体感気温で定義される「暑い日」「寒い日」に売れ行きが伸びる商品カテゴリの特定を行った。その結果、本研究で提案した体感気温の有効性を示すことができた。

今後の課題としては、温度感覚に関わる単語群の抽出、体感気温の算出に用いる気象要素の検討、需要予測に有用な体感気温の定量化方法の検討などが挙げられる。

参考文献

- [1] A. Missenard. "On Thermally Equivalent Environments," JIHVE, Vol. 27, pp. 231–237, 1959.
- [2] E. C. Thom. "The Discomfort Index," Weatherwise, Vol. 12, No. 2, pp. 57–61, 1959.
- [3] 久保拓弥, "データ解析のための統計モデリング入門," 岩波書店, 2012.
- [4] F. M. Tseng, H. C. Yu, and G. H. Tzeng. "Applied Hybrid Grey Model to Forecast Seasonal Time Series," *Technological Forecasting and Social Change*, Vol. 67, No. 2, pp. 291–302, 2001.
- [5] L. Breiman "Random Forests," *Machine Learning*, Vol. 45, pp. 5–32, 2001.